

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)<sup>a</sup>

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
16 August 2001 (16.08.2001)

PCT

(10) International Publication Number  
**WO 01/59133 A1**

(51) International Patent Classification<sup>7</sup>: **C12N 15/57**,  
9/64, 15/62, A61K 38/48, C12Q 1/68, C12N 5/10, G01N  
33/50, C12Q 1/37, C07K 16/40 // A61P 35/00

**HENTSCH, Bernd** [DE/DE]; Kaupstrasse 48, 64289  
Darmstadt (DE). **HOHEISEL, Jörg** [DE/DE]; Lempen-  
seite 18-5, 69168 Wiesloch (DE). **FROHME, Marcus**  
[DE/DE]; Danziger Strasse 6, 68535 Edingen (DE).

(21) International Application Number: PCT/EP01/01525

(74) Common Representative: **MERCK PATENT GMBH**;  
Frankfurter Strasse 250, 64293 Darmstadt (DE).

(22) International Filing Date: 12 February 2001 (12.02.2001)

(25) Filing Language: English

(81) Designated States (*national*): CA, JP, US.

(26) Publication Language: English

(84) Designated States (*regional*): European patent (AT, BE,  
CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC,  
NL, PT, SE, TR).

(30) Priority Data:  
00102955.2 14 February 2000 (14.02.2000) EP

**Published:**

- with international search report
- before the expiration of the time limit for amending the  
claims and to be republished in the event of receipt of  
amendments

(71) Applicant (*for all designated States except US*): **MERCK  
PATENT GMBH** [DE/DE]; Frankfurter Strasse 250,  
64293 Darmstadt (DE).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **DÜCKER, Klaus**  
[DE/DE]; Etteskerstrasse 5, 64291 Darmstadt (DE).

*For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*



**WO 01/59133 A1**

(54) Title: LARYNX CARCINOMA-ASSOCIATED PROTEIN LARCAP-1

(57) Abstract: Larynx carcinoma associated protein-1 (LarCAP-1) polypeptides and polynucleotides and methods for producing such polypeptides by recombinant techniques are disclosed. Also disclosed are methods for utilizing LarCAP-1 polypeptides and polynucleotides in diagnostic assays.

## LARYNX CARCINOMA-ASSOCIATED PROTEIN LARCAP-1

**Field of the Invention**

This invention relates to newly identified polypeptides and polynucleotides encoding such polypeptides sometimes hereinafter referred to as „larynx carcinoma associated protein-1 (LarCAP-1)“, to their use in diagnosis and in identifying compounds that may be agonists, antagonists that are potentially useful in therapy, and to production of such polypeptides and polynucleotides.

**10 Background of the Invention**

The drug discovery process is currently undergoing a fundamental revolution as it embraces "functional genomics", that is, high throughput genome- or gene-based biology. This approach as a means to identify genes and gene products as therapeutic targets is rapidly superceding earlier approaches based on "positional cloning". A phenotype, that is a biological function or genetic disease, would be identified and this would then be tracked back to the responsible gene, based on its genetic map position.

Functional genomics relies heavily on high-throughput DNA sequencing technologies and the various tools of bioinformatics to identify gene sequences of potential interest from the many molecular biology databases now available. There is a continuing need to identify and characterise further genes and their related polypeptides/proteins, as targets for drug discovery.

**Summary of the Invention**

25 The present invention relates to larynx carcinoma associated protein-1 (LarCAP-1), in particular larynx carcinoma associated protein-1 (LarCAP-1) polypeptides and larynx carcinoma associated protein-1 (LarCAP-1) polynucleotides, recombinant materials and methods for their production. Such polypeptides and polynucleotides are of interest in relation to methods of treatment of certain diseases, including, but not limited to, carcinomas (esp. but not limited to larynx carcinoma), metastasis, arthritis, osteoporosis, immune disorders, stroke, ischemia, autoimmune diseases, angiogenesis, skin disorders and organ malformations, esp. but not limited to heart hypertrophy, hereinafter referred to as " diseases of the

invention". In a further aspect, the invention relates to methods for identifying agonists and antagonists (e.g., inhibitors) using the materials provided by the invention, and treating conditions associated with larynx carcinoma associated protein-1 (LarCAP-1) imbalance with the identified compounds. In a still further aspect, the invention relates to diagnostic assays for detecting diseases associated with inappropriate larynx carcinoma associated protein-1 (LarCAP-1) activity or levels.

### Description of the Invention

- 10 In a first aspect, the present invention relates to larynx carcinoma associated protein-1 (LarCAP-1) polypeptides. Such polypeptides include:
- (a) a polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1;
  - (b) a polypeptide comprising a polypeptide sequence having at least 95%, 96%,  
15 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;
  - (c) a polypeptide comprising the polypeptide sequence of SEQ ID NO:2;
  - (d) a polypeptide having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;
  - (e) the polypeptide sequence of SEQ ID NO:2; and
  - 20 (f) a polypeptide having or comprising a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polypeptide sequence of SEQ ID NO:2;
  - (g) fragments and variants of such polypeptides in (a) to (f).

Polypeptides of the present invention are believed to be members of the protease family of polypeptides. Larynx-cancer associated protein-1 (LarCAP-1) has  
25 originally been identified in a screen searching for genes upregulated in larynx cancer. It encodes a novel protein with close sequence homology to a class of zinc metalloproteases known as A Disintegrin And Metalloproteinase with Thrombospondin motifs, ADAM-TS (Hurskainen T.L., *et al.*, J. Biol. Chem. 274,  
30 25555-25563, 1999). Only for two of the eight meanwhile known members of the ADAM-TS family an *in vivo* function has been identified so far. ADAM-TS1 is selectively expressed in a mouse cachexigenic colon cancer model and further examination showed that it is generally upregulated upon inflammation (Kuno, K.

*et al.*, J. Biol. Chem. 272, 556-562, 1997). Ectopic expression shows a clear extracellular matrix association, most likely caused by interaction with heparin or heparan-sulfate proteoglycans. ADAM-TS-2, also known as procollagen-1 N-proteinase or PCINP is involved in procollagen I (and perhaps collagen XIV) processing and mutations in this gene are known to cause Ehlers-Danlos syndrome VIIC (Smith, T.L. *et al.*, Am. J. Hum. Genet. 51, 235-244, 1992; Lapiere, C.M. and Nusgens, B.V., Arch. Dermatol. 129, 1316-1319, 1993). Because the herewith disclosed novel member of the ADAM-TS family shows highest sequence conservation to ADAM-TS2 (PCINP) and ADAM-TS3 (KIAA0366), it is possible that it possesses a similar function to them, and could therefore be involved in the maturation or degradation of the extracellular matrix (ECM) and/or processing or release of ECM-associated proteins, which often function in various signalling pathways. This feature is highly important during organ growth, inflammatory processes and cell migration (including metastasis) and supports the assumption that this gene plays an important role in -including but not limited to- larynx cancer.

The biological properties of the larynx carcinoma associated protein-1 (LarCAP-1) are hereinafter referred to as "biological activity of larynx carcinoma associated protein-1 (LarCAP-1)" or "larynx carcinoma associated protein-1 (LarCAP-1) activity". Preferably, a polypeptide of the present invention exhibits at least one biological activity of larynx carcinoma associated protein-1 (LarCAP-1).

Polypeptides of the present invention also includes variants of the aforementioned polypeptides, including all allelic forms and splice variants. Such polypeptides vary from the reference polypeptide by insertions, deletions, and substitutions that may be conservative or non-conservative, or any combination thereof. Particularly preferred variants are those in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5, from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acids are inserted, substituted, or deleted, in any combination.

Preferred fragments of polypeptides of the present invention include an isolated polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids from the amino acid sequence of SEQ ID NO: 2, or an isolated polypeptide comprising an amino acid sequence having at least 30, 50 or 100 contiguous amino acids truncated or deleted from the amino acid sequence

of SEQ ID NO: 2. Preferred fragments are biologically active fragments that mediate the biological activity of larynx carcinoma associated protein-1 (LarCAP-1), including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also preferred are those fragments that are antigenic or immunogenic in an animal, especially in a human.

Fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention. The polypeptides of the present invention may be in the form of the "mature" protein or may be a part of a larger protein such as a precursor or a fusion protein. It is often advantageous to include an additional amino acid sequence that contains secretory or leader sequences, pro-

sequences, sequences that aid in purification, for instance multiple histidine residues, or an additional sequence for stability during recombinant production. Polypeptides of the present invention can be prepared in any suitable manner, for instance by isolation from naturally occurring sources, from genetically engineered host cells comprising expression systems (*vide infra*) or by chemical synthesis, using for instance automated peptide synthesizers, or a combination of such methods.. Means for preparing such polypeptides are well understood in the art.

20

In a further aspect, the present invention relates to larynx carcinoma associated protein-1 (LarCAP-1) polynucleotides. Such polynucleotides include:

- (a) a polynucleotide comprising a polynucleotide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide sequence of SEQ ID NO:1;
- 25 (b) a polynucleotide comprising the polynucleotide of SEQ ID NO:1;
- (c) a polynucleotide having at least 95%, 96%, 97%, 98%, or 99% identity to the polynucleotide of SEQ ID NO:1;
- (d) the polynucleotide of SEQ ID NO:1;
- (e) a polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;
- 30 (f) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2;

- (g) a polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95%, 96%, 97%, 98%, or 99% identity to the polypeptide sequence of SEQ ID NO:2;
- (h) a polynucleotide encoding the polypeptide of SEQ ID NO:2;
- 5 (i) a polynucleotide having or comprising a polynucleotide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99 compared to the polynucleotide sequence of SEQ ID NO:1;
- (j) a polynucleotide having or comprising a polynucleotide sequence encoding a polypeptide sequence that has an Identity Index of 0.95, 0.96, 0.97, 0.98, or 0.99
- 10 compared to the polypeptide sequence of SEQ ID NO:2; and
- polynucleotides that are fragments and variants of the above mentioned polynucleotides or that are complementary to above mentioned polynucleotides, over the entire length thereof.

Preferred fragments of polynucleotides of the present invention include an isolated

15 polynucleotide comprising an nucleotide sequence having at least 15, 30, 50 or 100 contiguous nucleotides from the sequence of SEQ ID NO: 1, or an isolated polynucleotide comprising an sequence having at least 30, 50 or 100 contiguous nucleotides truncated or deleted from the sequence of SEQ ID NO: 1.

Preferred variants of polynucleotides of the present invention include splice

20 variants, allelic variants, and polymorphisms, including polynucleotides having one or more single nucleotide polymorphisms (SNPs).

Polynucleotides of the present invention also include polynucleotides encoding polypeptide variants that comprise the amino acid sequence of SEQ ID NO:2 and in which several, for instance from 50 to 30, from 30 to 20, from 20 to 10, from 10 to 5,

25 from 5 to 3, from 3 to 2, from 2 to 1 or 1 amino acid residues are substituted, deleted or added, in any combination.

In a further aspect, the present invention provides polynucleotides that are RNA transcripts of the DNA sequences of the present invention. Accordingly, there is provided an RNA polynucleotide that:

- 30 (a) comprises an RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2;
- (b) is the RNA transcript of the DNA sequence encoding the polypeptide of SEQ ID NO:2;

- 6 -

(c) comprises an RNA transcript of the DNA sequence of SEQ ID NO:1; or  
(d) is the RNA transcript of the DNA sequence of SEQ ID NO:1;  
and RNA polynucleotides that are complementary thereto.

- 5 The polynucleotide sequence of SEQ ID NO:1 shows homology with HSAJ3125 (Colige A.C. et al., unpublished) and AB002364 (Nagase, T. et al., unpublished). The polynucleotide sequence of SEQ ID NO:1 is a cDNA sequence that encodes the polypeptide of SEQ ID NO:2. The polynucleotide sequence encoding the polypeptide of SEQ ID NO:2 may be identical to the polypeptide encoding  
10 sequence of SEQ ID NO:1 or it may be a sequence other than SEQ ID NO:1, which, as a result of the redundancy (degeneracy) of the genetic code, also encodes the polypeptide of SEQ ID NO:2. The polypeptide of the SEQ ID NO:2 is related to other proteins of the protease family, having homology and/or structural similarity with GI-3928000 (Colige, A.C. et al., unpublished) and GI-2224673  
15 (Nagase, T. et al., unpublished).

Preferred polypeptides and polynucleotides of the present invention are expected to have, *inter alia*, similar biological functions/properties to their homologous polypeptides and polynucleotides. Furthermore, preferred polypeptides and polynucleotides of the present invention have at least one larynx carcinoma  
20 associated protein-1 (LarCAP-1) activity.

Polynucleotides of the present invention may be obtained using standard cloning and screening techniques from a cDNA library derived from mRNA in cells of human larynx carcinoma, heart, stomach, colon, pancreas, foreskin, whole embryo ,  
25 (see for instance, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

- 30 When polynucleotides of the present invention are used for the recombinant production of polypeptides of the present invention, the polynucleotide may include the coding sequence for the mature polypeptide, by itself, or the coding sequence for the mature polypeptide in reading frame with other coding sequences,

- such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro-protein sequence, or other fusion peptide portions. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, Proc Natl Acad Sci USA (1989) 86:821-824, or is an HA tag. The polynucleotide may also contain non-coding 5' and 3' sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.
- 10 Polynucleotides that are identical, or have sufficient identity to a polynucleotide sequence of SEQ ID NO:1, may be used as hybridization probes for cDNA and genomic DNA or as primers for a nucleic acid amplification reaction (for instance, PCR). Such probes and primers may be used to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes (including genes encoding paralogs from human sources and orthologs and paralogs from species other than human) that have a high sequence similarity to SEQ ID NO:1, typically at least 95% identity. Preferred probes and primers will generally comprise at least 15 nucleotides, preferably, at least 30 nucleotides and may have at least 50, if not at least 100 nucleotides. Particularly preferred probes will have between 30 and 50 nucleotides. Particularly preferred primers will have between 20 and 25 nucleotides.
- 25 A polynucleotide encoding a polypeptide of the present invention, including homologs from species other than human, may be obtained by a process comprising the steps of screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof, preferably of at least 15 nucleotides; and isolating full-length cDNA and genomic clones containing said polynucleotide sequence. Such hybridization techniques are well known to the skilled artisan. Preferred stringent hybridization conditions include overnight incubation at 42°C in a solution comprising: 50% formamide, 5xSSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10 % dextran sulfate, and 20 microgram/ml denatured, sheared salmon sperm DNA; followed by washing the filters in 0.1x SSC at about 65°C. Thus the present invention also includes isolated polynucleotides,
- 30



preferably with a nucleotide sequence of at least 100, obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO:1 or a fragment thereof, preferably of at least 15 nucleotides.

- 5 The skilled artisan will appreciate that, in many cases, an isolated cDNA sequence will be incomplete, in that the region coding for the polypeptide does not extend all the way through to the 5' terminus. This is a consequence of reverse transcriptase, an enzyme with inherently low "processivity" (a measure of the ability of the enzyme to remain attached to the template during the polymerisation reaction), failing to complete a DNA copy of the mRNA template
- 10 during first strand cDNA synthesis.

There are several methods available and well known to those skilled in the art to obtain full-length cDNAs, or extend short cDNAs, for example those based on the method of Rapid Amplification of cDNA ends (RACE) (see, for example, Frohman

15 et al., Proc Nat Acad Sci USA 85, 8998-9002, 1988). Recent modifications of the technique, exemplified by the Marathon (trade mark) technology (Clontech Laboratories Inc.) for example, have significantly simplified the search for longer cDNAs. In the Marathon (trade mark) technology, cDNAs have been prepared from mRNA extracted from a chosen tissue and an 'adaptor' sequence ligated

20 onto each end. Nucleic acid amplification (PCR) is then carried out to amplify the "missing" 5' end of the cDNA using a combination of gene specific and adaptor specific oligonucleotide primers. The PCR reaction is then repeated using 'nested' primers, that is, primers designed to anneal within the amplified product (typically an adaptor specific primer that anneals further 3' in the adaptor

25 sequence and a gene specific primer that anneals further 5' in the known gene sequence). The products of this reaction can then be analysed by DNA sequencing and a full-length cDNA constructed either by joining the product directly to the existing cDNA to give a complete sequence, or carrying out a separate full-length PCR using the new sequence information for the design of

30 the 5' primer.

Recombinant polypeptides of the present invention may be prepared by processes well known in the art from genetically engineered host cells comprising expression

systems. Accordingly, in a further aspect, the present invention relates to expression systems comprising a polynucleotide or polynucleotides of the present invention, to host cells which are genetically engineered with such expression systems and to the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Polynucleotides may be introduced into host cells by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology (1986) and Sambrook et al. (*ibid*). Preferred methods of introducing polynucleotides into host cells include, for instance, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as *Streptococci*, *Staphylococci*, *E. coli*, *Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, HEK 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used, for instance, chromosomal, episomal and virus-derived systems, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector that is able to maintain, propagate or express a polynucleotide to produce a polypeptide in a host may be used. The appropriate polynucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook et al., (*ibid*).

Appropriate secretion signals may be incorporated into the desired polypeptide to allow secretion of the translated protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment. These signals may be endogenous to the polypeptide or they may be heterologous signals.

- 5 If a polypeptide of the present invention is to be expressed for use in screening assays, it is generally preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If the polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide. If produced  
10 intracellularly, the cells must first be lysed before the polypeptide is recovered.

Polypeptides of the present invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity  
15 chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during intracellular synthesis, isolation and/or purification.

- 20 Polynucleotides of the present invention may be used as diagnostic reagents, through detecting mutations in the associated gene. Detection of a mutated form of the gene characterised by the polynucleotide of SEQ ID NO:1 in the cDNA or genomic sequence and which is associated with a dysfunction will provide a diagnostic tool that can add to, or define, a diagnosis of a disease, or susceptibility  
25 to a disease, which results from under-expression, over-expression or altered spatial or temporal expression of the gene. Individuals carrying mutations in the gene may be detected at the DNA level by a variety of techniques well known in the art.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from  
30 blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA may be used directly for detection or it may be amplified enzymatically by using PCR, preferably RT-PCR, or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be

detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled larynx carcinoma associated protein-1 (LarCAP-1) nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by  
5 RNase digestion or by differences in melting temperatures. DNA sequence difference may also be detected by alterations in the electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing (see, for instance, Myers *et al.*, Science (1985) 230:1242). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as  
10 RNase and S1 protection or the chemical cleavage method (see Cotton *et al.*, Proc Natl Acad Sci USA (1985) 85: 4397-4401).

An array of oligonucleotides probes comprising larynx carcinoma associated protein-1 (LarCAP-1) polynucleotide sequence or fragments thereof can be constructed to conduct efficient screening of e.g., genetic mutations. Such arrays  
15 are preferably high density arrays or grids. Array technology methods are well known and have general applicability and can be used to address a variety of questions in molecular genetics including gene expression, genetic linkage, and genetic variability, see, for example, M.Chee *et al.*, Science, 274, 610-613 (1996) and other references cited therein.

20 Detection of abnormally decreased or increased levels of polypeptide or mRNA expression may also be used for diagnosing or determining susceptibility of a subject to a disease of the invention. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, nucleic acid amplification,  
25 for instance PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as a polypeptide of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA  
30 assays.

Thus in another aspect, the present invention relates to a diagnostic kit comprising:

- 12 -

(a) a polynucleotide of the present invention, preferably the nucleotide sequence of SEQ ID NO: 1, or a fragment or an RNA transcript thereof;

(b) a nucleotide sequence complementary to that of (a);

(c) a polypeptide of the present invention, preferably the polypeptide of SEQ ID NO:2 or a fragment thereof; or

(d) an antibody to a polypeptide of the present invention, preferably to the polypeptide of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component. Such a kit will be of use in diagnosing a disease or susceptibility to a disease, particularly diseases of the invention, amongst others.

The polynucleotide sequences of the present invention are valuable for chromosome localisation studies. The sequence is specifically targeted to, and can hybridize with, a particular location on an individual human chromosome. The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with gene associated disease. Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found in, for example, V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (co-inheritance of physically adjacent genes). Precise human chromosomal localisations for a genomic sequence (gene fragment etc.) can be determined using Radiation Hybrid (RH) Mapping (Walter, M. Spillet, D., Thomas, P., Weissenbach, J., and Goodfellow, P., (1994) A method for constructing radiation hybrid maps of whole genomes, Nature Genetics 7, 22-28). A number of RH panels are available from Research Genetics (Huntsville, AL, USA) e.g. the GeneBridge4 RH panel (Hum Mol Genet 1996 Mar;5(3):339-46 A radiation hybrid map of the human genome. Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillet D, Muselet D, Prud'Homme JF, Dib C, Auffray C, Morissette J, Weissenbach J, Goodfellow PN). To determine the chromosomal location of a gene using this panel, 93 PCRs are performed using

primers designed from the gene of interest on RH DNAs. Each of these DNAs contains random human genomic fragments maintained in a hamster background (human / hamster hybrid cell lines). These PCRs result in 93 scores indicating the presence or absence of the PCR product of the gene of interest. These  
5 scores are compared with scores created using PCR products from genomic sequences of known location. This comparison is conducted at <http://www.genome.wi.mit.edu/>. The gene of the present invention maps to human chromosome \*LOCATION.

10 The polynucleotide sequences of the present invention are also valuable tools for tissue expression studies. Such studies allow the determination of expression patterns of polynucleotides of the present invention which may give an indication as to the expression patterns of the encoded polypeptides in tissues, by detecting the mRNAs that encode them. The techniques used are well known in the art and  
15 include in situ hybridisation techniques to clones arrayed on a grid, such as cDNA microarray hybridisation (Schena *et al*, Science, 270, 467-470, 1995 and Shalon *et al*, Genome Res, 6, 639-645, 1996) and nucleotide amplification techniques such as PCR. A preferred method uses the TAQMAN (Trade mark) technology available from Perkin Elmer. Results from these studies can provide an indication of the  
20 normal function of the polypeptide in the organism. In addition, comparative studies of the normal expression pattern of mRNAs with that of mRNAs encoded by an alternative form of the same gene (for example, one having an alteration in polypeptide coding potential or a regulatory mutation) can provide valuable insights into the role of the polypeptides of the present invention, or that of inappropriate  
25 expression thereof in disease. Such inappropriate expression may be of a temporal, spatial or simply quantitative nature.

The polypeptides of the present invention are expressed in larynx carcinoma, heart, stomach, colon, pancreas, foreskin, whole embryo .

30 A further aspect of the present invention relates to antibodies. The polypeptides of the invention or their fragments, or cells expressing them, can be used as immunogens to produce antibodies that are immunospecific for polypeptides of the present invention. The term "immunospecific" means that the antibodies have

substantially greater affinity for the polypeptides of the invention than their affinity for other related polypeptides in the prior art.

Antibodies generated against polypeptides of the present invention may be obtained by administering the polypeptides or epitope-bearing fragments, or cells to  
5 an animal, preferably a non-human animal, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature* (1975) 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al.*, *Immunology Today* (1983) 4:72) and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies, such as those described in U.S. Patent No. 4,946,778, can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms,  
15 including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography. Antibodies against polypeptides of the present invention may also be employed to treat diseases of the invention, amongst others.

20

Polypeptides and polynucleotides of the present invention may also be used as vaccines. Accordingly, in a further aspect, the present invention relates to a method for inducing an immunological response in a mammal that comprises inoculating the mammal with a polypeptide of the present invention, adequate to  
25 produce antibody and/or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said animal from disease, whether that disease is already established within the individual or not. An immunological response in a mammal may also be induced by a method comprises delivering a polypeptide of the present invention *via* a vector directing  
30 expression of the polynucleotide and coding for the polypeptide *in vivo* in order to induce such an immunological response to produce antibody to protect said animal from diseases of the invention. One way of administering the vector is by accelerating it into the desired cells as a coating on particles or otherwise. Such

- 15 -

nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid. For use as a vaccine, a polypeptide or a nucleic acid vector will be normally provided as a vaccine formulation (composition). The formulation may further comprise a suitable carrier. Since a polypeptide may be broken  
5 down in the stomach, it is preferably administered parenterally (for instance, subcutaneous, intramuscular, intravenous, or intradermal injection). Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions that may contain anti-oxidants, buffers, bacteriostats and solutes that render the formulation isotonic with the blood of the recipient; and  
10 aqueous and non-aqueous sterile suspensions that may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include  
15 adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

20 Polypeptides of the present invention have one or more biological functions that are of relevance in one or more disease states, in particular the diseases of the invention hereinbefore mentioned. It is therefore useful to identify compounds that stimulate or inhibit the function or level of the polypeptide. Accordingly, in a further aspect, the present invention provides for a method of screening compounds  
25 to identify those that stimulate or inhibit the function or level of the polypeptide. Such methods identify agonists or antagonists that may be employed for therapeutic and prophylactic purposes for such diseases of the invention as hereinbefore mentioned. Compounds may be identified from a variety of sources, for example, cells, cell-free preparations, chemical libraries, collections of chemical  
30 compounds, and natural product mixtures. Such agonists or antagonists so-identified may be natural or modified substrates, ligands, receptors, enzymes, etc., as the case may be, of the polypeptide; a structural or functional mimetic thereof



- 16 -

(see Coligan *et al.*, Current Protocols in Immunology 1(2):Chapter 5 (1991)) or a small molecule.

The screening method may simply measure the binding of a candidate compound to the polypeptide, or to cells or membranes bearing the polypeptide, or a fusion protein thereof, by means of a label directly or indirectly associated with the candidate compound. Alternatively, the screening method may involve measuring or detecting (qualitatively or quantitatively) the competitive binding of a candidate compound to the polypeptide against a labeled competitor (e.g. agonist or antagonist). Further, these screening methods may test whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells bearing the polypeptide. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Further, the screening methods may simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide of the present invention, to form a mixture, measuring a larynx carcinoma associated protein-1 (LarCAP-1) activity in the mixture, and comparing the larynx carcinoma associated protein-1 (LarCAP-1) activity of the mixture to a control mixture which contains no candidate compound.

Polypeptides of the present invention may be employed in conventional low capacity screening methods and also in high-throughput screening (HTS) formats. Such HTS formats include not only the well-established use of 96- and, more recently, 384-well micotiter plates but also emerging methods such as the nanowell method described by Schullek *et al*, Anal Biochem., 246, 20-29, (1997).

Fusion proteins, such as those made from Fc portion and larynx carcinoma associated protein-1 (LarCAP-1) polypeptide, as hereinbefore described, can also be used for high-throughput screening assays to identify antagonists for the polypeptide of the present invention (see D. Bennett *et al.*, J Mol Recognition, 8:52-58 (1995); and K. Johanson *et al.*, J Biol Chem, 270(16):9459-9471 (1995)).

30

\*Screening techniques

- 17 -

The polynucleotides, polypeptides and antibodies to the polypeptide of the present invention may also be used to configure screening methods for detecting the effect of added compounds on the production of mRNA and polypeptide in cells. For example, an ELISA assay may be constructed for measuring secreted or cell  
5 associated levels of polypeptide using monoclonal and polyclonal antibodies by standard methods known in the art. This can be used to discover agents that may inhibit or enhance the production of polypeptide (also called antagonist or agonist, respectively) from suitably manipulated cells or tissues.

A polypeptide of the present invention may be used to identify membrane bound  
10 or soluble receptors, if any, through standard receptor binding techniques known in the art. These include, but are not limited to, ligand binding and crosslinking assays in which the polypeptide is labeled with a radioactive isotope (for instance,  $^{125}\text{I}$ ), chemically modified (for instance, biotinylated), or fused to a peptide sequence suitable for detection or purification, and incubated with a source of the  
15 putative receptor (cells, cell membranes, cell supernatants, tissue extracts, bodily fluids). Other methods include biophysical techniques such as surface plasmon resonance and spectroscopy. These screening methods may also be used to identify agonists and antagonists of the polypeptide that compete with the binding of the polypeptide to its receptors, if any. Standard methods for conducting such  
20 assays are well understood in the art.

Examples of antagonists of polypeptides of the present invention include antibodies or, in some cases, oligonucleotides or proteins that are closely related to the ligands, substrates, receptors, enzymes, etc., as the case may be, of the polypeptide, e.g., a fragment of the ligands, substrates, receptors, enzymes, etc.; or  
25 a small molecule that bind to the polypeptide of the present invention but do not elicit a response, so that the activity of the polypeptide is prevented.

Screening methods may also involve the use of transgenic technology and larynx carcinoma associated protein-1 (LarCAP-1) gene. The art of constructing transgenic animals is well established. For example, the larynx carcinoma  
30 associated protein-1 (LarCAP-1) gene may be introduced through microinjection into the male pronucleus of fertilized oocytes, retroviral transfer into pre- or post-implantation embryos, or injection of genetically modified, such as by electroporation, embryonic stem cells into host blastocysts. Particularly useful

transgenic animals are so-called "knock-in" animals in which an animal gene is replaced by the human equivalent within the genome of that animal. Knock-in transgenic animals are useful in the drug discovery process, for target validation, where the compound is specific for the human target. Other useful transgenic animals are so-called "knock-out" animals in which the expression of the animal ortholog of a polypeptide of the present invention and encoded by an endogenous DNA sequence in a cell is partially or completely annulled. The gene knock-out may be targeted to specific cells or tissues, may occur only in certain cells or tissues as a consequence of the limitations of the technology, or may occur in all, or substantially all, cells in the animal. Transgenic animal technology also offers a whole animal expression-cloning system in which introduced genes are expressed to give large amounts of polypeptides of the present invention

Screening kits for use in the above described methods form a further aspect of the present invention. Such screening kits comprise:

- (a) a polypeptide of the present invention;
- (b) a recombinant cell expressing a polypeptide of the present invention;
- (c) a cell membrane expressing a polypeptide of the present invention; or
- (d) an antibody to a polypeptide of the present invention;

which polypeptide is preferably that of SEQ ID NO:2.

It will be appreciated that in any such kit, (a), (b), (c) or (d) may comprise a substantial component.

### Glossary

The following definitions are provided to facilitate understanding of certain terms used frequently hereinbefore.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an

Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a

living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Moreover, a polynucleotide or polypeptide that is introduced into an organism by transformation, genetic manipulation or by any other recombinant method is "isolated" even if it is still present in said organism, which organism may be living or non-living.

"Polynucleotide" generally refers to any polyribonucleotide (RNA) or polydeoxribonucleotide (DNA), which may be unmodified or modified RNA or DNA. "Polynucleotides" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term "polynucleotide" also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications may be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any polypeptide comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques that are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications may occur anywhere in a

polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present to the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, biotinylation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination (see, for instance, *Proteins - Structure and Molecular Properties*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993; Wold, F., *Post-translational Protein Modifications: Perspectives and Prospects*, 1-12, in *Post-translational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Meth Enzymol*, 182, 626-646, 1990, and Rattan *et al.*, "Protein Synthesis: Post-translational Modifications and Aging", *Ann NY Acad Sci*, 663, 48-62, 1992).

"Fragment" of a polypeptide sequence refers to a polypeptide sequence that is shorter than the reference sequence but that retains essentially the same biological function or activity as the reference polypeptide. "Fragment" of a polynucleotide sequence refers to a polynucleotide sequence that is shorter than the reference sequence of SEQ ID NO:1..

"Variant" refers to a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide, but retains the essential properties thereof. A typical variant of a polynucleotide differs in nucleotide sequence from the

reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from the reference polypeptide. Generally, alterations are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, insertions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. Typical conservative substitutions include Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe and Tyr. A variant of a polynucleotide or polypeptide may be naturally occurring such as an allele, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis. Also included as variants are polypeptides having one or more post-translational modifications, for instance glycosylation, phosphorylation, methylation, ADP ribosylation and the like. Embodiments include methylation of the N-terminal amino acid, phosphorylations of serines and threonines and modification of C-terminal glycines.

"Allele" refers to one of two or more alternative forms of a gene occurring at a given locus in the genome.

"Polymorphism" refers to a variation in nucleotide sequence (and/or encoded polypeptide sequence, if relevant) at a given position in the genome within a population.

"Single Nucleotide Polymorphism" (SNP) refers to the occurrence of nucleotide variability at a single nucleotide position in the genome, within a population. An SNP may occur within a gene or within intergenic regions of the genome. SNPs can be assayed using Allele Specific Amplification (ASA). For the process at least 3 primers are required. A common primer is used in reverse complement to the polymorphism being assayed. This common primer can be between 50 and 1500 bps from the polymorphic base. The other two (or more) primers are

- 22 -

identical to each other except that the final 3' base wobbles to match one of the two (or more) alleles that make up the polymorphism. Two (or more) PCR reactions are then conducted on sample DNA, each using the common primer and one of the Allele Specific Primers.

- 5 "Splice Variant" as used herein refers to cDNA molecules produced from RNA molecules initially transcribed from the same genomic DNA sequence but which have undergone alternative RNA splicing. Alternative RNA splicing occurs when a primary RNA transcript undergoes splicing, generally for the removal of introns, which results in the production of more than one mRNA molecule each of that  
10 may encode different amino acid sequences. The term splice variant also refers to the proteins encoded by the above cDNA molecules.

"Identity" reflects a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, determined by comparing the sequences. In general, identity refers to an exact nucleotide to nucleotide or amino acid to  
15 amino acid correspondence of the two polynucleotide or two polypeptide sequences, respectively, over the length of the sequences being compared.

"% Identity" - For sequences where there is not an exact correspondence, a "% identity" may be determined. In general, the two sequences to be compared are aligned to give a maximum correlation between the sequences. This may include  
20 inserting "gaps" in either one or both sequences, to enhance the degree of alignment. A % identity may be determined over the whole length of each of the sequences being compared (so-called global alignment), that is particularly suitable for sequences of the same or very similar length, or over shorter, defined lengths (so-called local alignment), that is more suitable for sequences of unequal  
25 length.

"Similarity" is a further, more sophisticated measure of the relationship between two polypeptide sequences. In general, "similarity" means a comparison between the amino acids of two polypeptide chains, on a residue by residue basis, taking into account not only exact correspondences between a between pairs of  
30 residues, one from each of the sequences being compared (as for identity) but also, where there is not an exact correspondence, whether, on an evolutionary basis, one residue is a likely substitute for the other. This likelihood has an

associated "score" from which the "% similarity" of the two sequences can then be determined.

Methods for comparing the identity and similarity of two or more sequences are well known in the art. Thus for instance, programs available in the Wisconsin Sequence Analysis Package, version 9.1 (Devereux J et al, Nucleic Acids Res, 12, 387-395, 1984, available from Genetics Computer Group, Madison, Wisconsin, USA), for example the programs BESTFIT and GAP, may be used to determine the % identity between two polynucleotides and the % identity and the % similarity between two polypeptide sequences. BESTFIT uses the "local  
10 homology" algorithm of Smith and Waterman (J Mol Biol, 147,195-197, 1981, Advances in Applied Mathematics, 2, 482-489, 1981) and finds the best single region of similarity between two sequences. BESTFIT is more suited to comparing two polynucleotide or two polypeptide sequences that are dissimilar in length, the program assuming that the shorter sequence represents a portion of  
15 the longer. In comparison, GAP aligns two sequences, finding a "maximum similarity", according to the algorithm of Neddleman and Wunsch (J Mol Biol, 48, 443-453, 1970). GAP is more suited to comparing sequences that are approximately the same length and an alignment is expected over the entire length. Preferably, the parameters "Gap Weight" and "Length Weight" used in  
20 each program are 50 and 3, for polynucleotide sequences and 12 and 4 for polypeptide sequences, respectively. Preferably, % identities and similarities are determined when the two sequences being compared are optimally aligned.

Other programs for determining identity and/or similarity between sequences are also known in the art, for instance the BLAST family of programs (Altschul S F et  
25 al, J Mol Biol, 215, 403-410, 1990, Altschul S F et al, Nucleic Acids Res., 25:389-3402, 1997, available from the National Center for Biotechnology Information (NCBI), Bethesda, Maryland, USA and accessible through the home page of the NCBI at [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) and FASTA (Pearson W R, Methods in Enzymology, 183, 63-99, 1990; Pearson W R and Lipman D J, Proc Nat Acad Sci  
30 USA, 85, 2444-2448, 1988, available as part of the Wisconsin Sequence Analysis Package).

Preferably, the BLOSUM62 amino acid substitution matrix (Henikoff S and Henikoff J G, Proc. Nat. Acad Sci. USA, 89, 10915-10919, 1992) is used in



polypeptide sequence comparisons including where nucleotide sequences are first translated into amino acid sequences before comparison.

Preferably, the program BESTFIT is used to determine the % identity of a query polynucleotide or a polypeptide sequence with respect to a reference  
5 polynucleotide or a polypeptide sequence, the query and the reference sequence being optimally aligned and the parameters of the program set at the default value,

"Identity Index" is a measure of sequence relatedness which may be used to compare a candidate sequence (polynucleotide or polypeptide) and a reference  
10 sequence. Thus, for instance, a candidate polynucleotide sequence having, for example, an Identity Index of 0.95 compared to a reference polynucleotide sequence is identical to the reference sequence except that the candidate polynucleotide sequence may include on average up to five differences per each 100 nucleotides of the reference sequence. Such differences are selected from  
15 the group consisting of at least one nucleotide deletion, substitution, including transition and transversion, or insertion. These differences may occur at the 5' or 3' terminal positions of the reference polynucleotide sequence or anywhere between these terminal positions, interspersed either individually among the nucleotides in the reference sequence or in one or more contiguous groups within  
20 the reference sequence. In other words, to obtain a polynucleotide sequence having an Identity Index of 0.95 compared to a reference polynucleotide sequence, an average of up to 5 in every 100 of the nucleotides of the in the reference sequence may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other  
25 values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

Similarly, for a polypeptide, a candidate polypeptide sequence having, for example, an Identity Index of 0.95 compared to a reference polypeptide sequence is identical to the reference sequence except that the polypeptide sequence may include an average of up to five differences per each 100 amino  
30 acids of the reference sequence. Such differences are selected from the group consisting of at least one amino acid deletion, substitution, including conservative and non-conservative substitution, or insertion. These differences may occur at the amino- or carboxy-terminal positions of the reference polypeptide sequence

or anywhere between these terminal positions, interspersed either individually among the amino acids in the reference sequence or in one or more contiguous groups within the reference sequence. In other words, to obtain a polypeptide sequence having an Identity Index of 0.95 compared to a reference polypeptide sequence, an average of up to 5 in every 100 of the amino acids in the reference sequence may be deleted, substituted or inserted, or any combination thereof, as hereinbefore described. The same applies *mutatis mutandis* for other values of the Identity Index, for instance 0.96, 0.97, 0.98 and 0.99.

The relationship between the number of nucleotide or amino acid differences and the Identity Index may be expressed in the following equation:

$$n_a \leq x_a - (x_a \cdot I),$$

in which:

$n_a$  is the number of nucleotide or amino acid differences,

$x_a$  is the total number of nucleotides or amino acids in SEQ ID NO:1 or SEQ ID

NO:2, respectively,

$I$  is the Identity Index ,

$\cdot$  is the symbol for the multiplication operator, and

in which any non-integer product of  $x_a$  and  $I$  is rounded down to the nearest integer prior to subtracting it from  $x_a$ .

"Homolog" is a generic term used in the art to indicate a polynucleotide or polypeptide sequence possessing a high degree of sequence relatedness to a reference sequence. Such relatedness may be quantified by determining the degree of identity and/or similarity between the two sequences as hereinbefore defined. Falling within this generic term are the terms "ortholog", and "paralog".

"Ortholog" refers to a polynucleotide or polypeptide that is the functional equivalent of the polynucleotide or polypeptide in another species. "Paralog" refers to a polynucleotide or polypeptide that within the same species which is functionally similar.

"Fusion protein" refers to a protein encoded by two, unrelated, fused genes or fragments thereof. Examples have been disclosed in US 5541087, 5726044. In the case of Fc-LarCAP-1, employing an immunoglobulin Fc region as a part of a fusion protein is advantageous for performing the functional expression of Fc-LarCAP-1 or fragments of LarCAP-1, to improve pharmacokinetic properties of

such a fusion protein when used for therapy and to generate a dimeric Fc-LarCAP-1. The Fc- LarCAP-1 DNA construct comprises in 5' to 3' direction, a secretion cassette, i.e. a signal sequence that triggers export from a mammalian cell, DNA encoding an immunoglobulin Fc region fragment, as a fusion partner,  
 5 and a DNA encoding Fc- LarCAP-1 or fragments thereof. In some uses it would be desirable to be able to alter the intrinsic functional properties (complement binding, Fc-Receptor binding) by mutating the functional Fc sides while leaving the rest of the fusion protein untouched or delete the Fc part completely after expression.

10

All publications and references, including but not limited to patents and patent applications, cited in this specification are herein incorporated by reference in their entirety as if each individual publication or reference were specifically and individually indicated to be incorporated by reference herein as being fully set  
 15 forth. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety in the manner described above for publications and references.

#### Figure legend:

**Figure 1.** LarCAP-1 is expressed in various human tissues and organs. The  
 20 labeling of bars in Figure 1 is as follows:

1-3	adult brain	20-21	fetal liver
4-5	fetal brain	22-24	pancreas
6-8	heart	25-26	spleen
9-10	skeletal muscle	27-28	kidney
11-12	small intestine	29-30	prostate
13-15	lung	31-32	uterus
16-17	fetal lung	33-34	placenta
18-19	adult liver	35-36	testis

## Further examples

### Detection of tissue and organe specific expression of LarCAP-1

Filter spotting, hybridization, data aquisition and normalization process Clones were obtained as either glycerol stock cultures or purified plasmid solutions.

5 Using standard PCR protocols and Qiagen HotStarTaq™ Master Mix, inserts of the respective plasmids were amplified using standard primers binding next to the insert within the plasmid backbones. Aliquots of the PCR reactions were individually checked for success by agarose gelelectrophoresis. PCR products were spotted without further purification in a twofold aqueaous dilution directly

10 onto Nylon membranes (Schleicher and Schuell NytranSuperCharge™). Spotting was performed with a Beckman Biomek 2000™ equipped with a 384- pin high density replica tool (HDRT) with flat ended pins of a diameter of 1.14 mm. Dry membranes were then crosslinked with 50 mJ using a BioRad GS GeneLinker™ UV device. After crosslinking, membranes were stored under dry conditions until

15 usage. Prehybridization and hybridization were performed in a temperature controlled hybridization oven equipped with rotating tubes with 15 ml each of Clontech ExpressHyb™ hybridization solutions at 50°C for 3 hrs and 16 hrs, respectively. Washings of the filters were performed successively with 0.8xSSC/0.1% SDS twice at 50°C for 20 min, 0.1xSSC/0.1% SDS at 50°C for 20 min,

20 0.1xSSC/0.1% SDS twice at 65°C for 40 min each. After the washing procedure, filters were semi-dried and wrapped in extremely thin polyethylen foil. Then a Fuji Phoshorimager Image Plate (IP) was exposed to the filters within a Fuji FLA3000™ exposition chamber. After 48- 72 hours IP plates were scanned with a Fuji Phosphorimager FLA3000™ device with a resolution of 50 µm. Using a

25 software package from Raytest (AIDA analyzer™), a grid was projected over the dots in the filter area, manually adjusted and fine adjusted using the respective software tool. Next, dot finding optimization and localized backgroud subtraction processing was performed. This generated a file with numerical data output for each spot position, corresponding over a wide range to the radioactivity present

30 at the respective positions. These data were used as data for all further calculations. To normalize filters, the arithmetic mean of all positions on a given filter was determined and the individual spot intensity divided by this mean, so

- 28 -

that in case all values would be equal, each spot would acquire a numerical value of „1“. These normalized signal intensities were used to compare different filters and are the data shown on all graphic chart displays.

### Generation of radioactive labeled probes

5 Probe labeling was performed in two different methods, either annotated as „mixed labeling“ or as „polydT- labeling“. RNA of different human tissues was obtained from Clontech. For polydT – labeling, an aliquot of 10 to 20 µg was derived from the suspension provided by Clontech and further processed by a sedimentation followed by washing with isopropanol. After drying at 50°C for  
10 30 min under agitation, the pellet was resuspended in purified water and 4 to 8 µg of total RNA was hybridized with 0.25 µg of poly dT primer [5' T<sub>12</sub>NNN3'] in a total volume of 21 µl. Primer annealing was performed at 65°C for 3 min followed by incubation in water at 0°C. Then a cocktail was added to each sample consisting of [per rxn] 4.2 µl purified water, 8 µl RT buffer from Promega enzyme M-MLV,  
15 0.8 µl mix of nucleotides dA, dT, dG [25 mM each in mixture], 5 µl alpha <sup>33</sup>[P] labeled dCTP [Amersham Redivue™ AA9905] and 1 µl [200 units] Promega M-MLV reverse transcriptase. Enzymatic labeling was performed under these dCTP- limited conditions for 25 min at 39°C followed by addition of a new cocktail consisting of [per rxn] 6.2 µl purified water, 2 µl M-MLV- buffer from Promega,  
20 0.8 µl dCTP [25 mM] and 1 µl [200 units] of Promega M-MLV reverse transcriptase. After incubation at 39°C for further 15 min the entire volume was Sephadex™ G25-column purified using Boehringer Quick Spin columns™ and an appropriate centrifugation device. The eluate was next denaturated at 95°C for 3 min in a thermoheater and chilled in water at 0°C. The entire probe volume was  
25 then added to 15 mls of preheated Clontech ExpressHyb™ hybridization solution and transferred to the prehybridized filters.

For the mixed labeling, RNA was obtained as aqueous solution of polyA enriched RNA from Clontech, representing various human tissues. For labeling, 0.5 µg of polyA enriched RNA was mixed with 0.25 µg of poly dT primer [5' T<sub>12</sub>NNN3'] and  
30 additional 0.5 µg of random primer [mostly N<sub>6</sub>, GibcoBRL] in a total volume of 21 µl. Further processing was as described for polydT- labeling.

**Claims**

1. A polypeptide selected from the group consisting of:
  - (a) a polypeptide encoded by a polynucleotide comprising the sequence of SEQ ID NO:1;
  - 5 (b) a polypeptide comprising a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;
  - c) a polypeptide having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;
  - d) the polypeptide sequence of SEQ ID NO:2 and
  - 10 (e) fragments and variants of such polypeptides in (a) to (d).
2. The polypeptide of claim 1 comprising the polypeptide sequence of SEQ ID NO:2.
- 15 3. The polypeptide of claim 1 which is the polypeptide sequence of SEQ ID NO:2.
4. A polynucleotide selected from the group consisting of:
  - (a) a polynucleotide comprising a polynucleotide sequence having at least 95% identity to the polynucleotide sequence of SEQ ID NO:1;
  - 20 (b) a polynucleotide having at least 95% identity to the polynucleotide of SEQ ID NO:1;
  - (c) a polynucleotide comprising a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;
  - 25 (d) a polynucleotide having a polynucleotide sequence encoding a polypeptide sequence having at least 95% identity to the polypeptide sequence of SEQ ID NO:2;

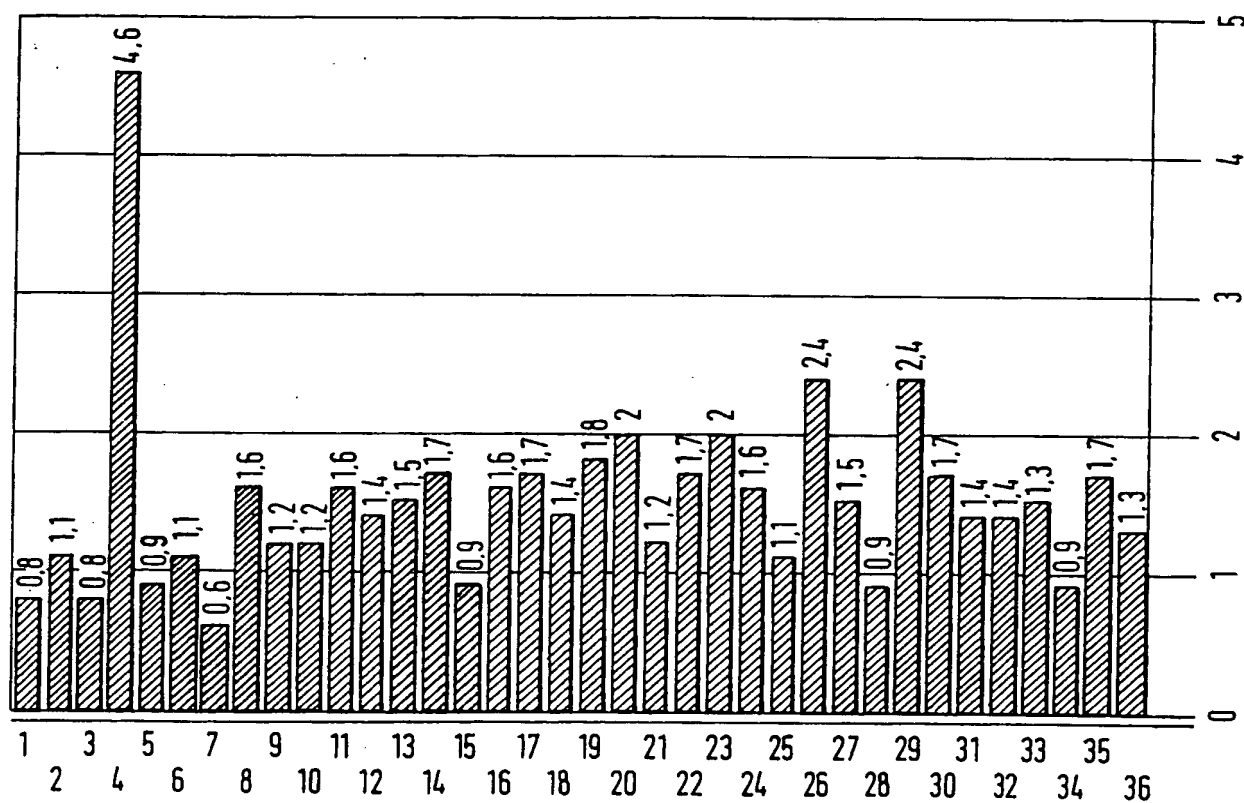
- (e) a polynucleotide with a nucleotide sequence of at least 100 nucleotides obtained by screening a library under stringent hybridization conditions with a labeled probe having the sequence of SEQ ID NO: 1 or a fragment thereof having at least 15 nucleotides;
- 5 (f) a polynucleotide which is the RNA equivalent of a polynucleotide of (a) to (e);
- (g) a polynucleotide sequence complementary to said polynucleotide of any one of (a) to (f), and
- (h) polynucleotides that are variants or fragments of the polynucleotides of any one of (a) to (g) or that are complementary to above mentioned polynucleotides,
- 10 over the entire length thereof.
5. A polynucleotide of claim 4 selected from the group consisting of:
- (a) a polynucleotide comprising the polynucleotide of SEQ ID NO:1;
- (b) the polynucleotide of SEQ ID NO:1;
- 15 (c) a polynucleotide comprising a polynucleotide sequence encoding the polypeptide of SEQ ID NO:2; and
- (d) a polynucleotide encoding the polypeptide of SEQ ID NO:2.
6. An expression system comprising a polynucleotide capable of producing a
- 20 polypeptide of any one of claim 1-3 when said expression vector is present in a compatible host cell.
7. A recombinant host cell comprising the expression vector of claim 6 or a membrane thereof expressing the polypeptide of any one of claim 1-3.
- 25
8. A process for producing a polypeptide of any one of claim 1-3 comprising the step of culturing a host cell as defined in claim 7 under conditions sufficient for

the production of said polypeptide and recovering the polypeptide from the culture medium.

9. A fusion protein consisting of the Immunoglobulin Fc-region and a polypeptide  
5 any one one of claims 1-3.
10. An antibody immunospecific for the polypeptide of any one of claims 1 to 3.
11. A method for screening to identify compounds that stimulate or inhibit the function or level of the polypeptide of any one of claim 1-3 comprising a method selected from the group consisting of:
- 10 (a) measuring or, detecting, quantitatively or qualitatively, the binding of a candidate compound to the polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof by means of a label directly or indirectly associated with the candidate compound;
- (b) measuring the competition of binding of a candidate compound to the  
15 polypeptide (or to the cells or membranes expressing the polypeptide) or a fusion protein thereof in the presence of a labeled competitor;
- (c) testing whether the candidate compound results in a signal generated by activation or inhibition of the polypeptide, using detection systems appropriate to the cells or cell membranes expressing the polypeptide;
- 20 (d) mixing a candidate compound with a solution containing a polypeptide of any one of claims 1-3, to form a mixture, measuring activity of the polypeptide in the mixture, and comparing the activity of the mixture to a control mixture which contains no candidate compound; or
- (e) detecting the effect of a candidate compound on the production of mRNA  
25 encoding said polypeptide or said polypeptide in cells, using for instance, an ELISA assay, and
- (f) producing said compound according to biotechnological or chemical standard techniques.



1/1



- 1 -

## SEQUENCE LISTING

5 <110> Merck Patent GmbH  
 <120> Tumor associated antigen  
 10 <130> dueck01ws  
 <140>  
 15 <141>  
 <160> 2  
 20 <170> PatentIn Ver. 2.1  
 <210> 1  
 <211> 3144  
 <212> DNA  
 25 <213> Homo sapiens  
 <220>  
 <221> CDS  
 <222> (1)..(3144)  
 30 <400> 1  
 atg cac gcg gct tct gct gcc cac agt ttc gtg ctt aca gcc tgg ctg  
 48  
 Met His Ala Ala Ser Ala Ala His Ser Phe Val Leu Thr Ala Trp Leu  
 35 1 5 10 15  
 cct cct tgg ctg gcc cag aag ggc tcc cag gtg ctg ctg ctg ctc caa  
 96  
 Pro Pro Trp Leu Ala Gln Lys Gly Ser Gln Val Leu Leu Leu Gln  
 40 20 25 30  
 ggt cct gga gta tgt ctc ctg gag gag tgc cca ctg ctg cct ccc cag  
 144  
 Gly Pro Gly Val Cys Leu Leu Glu Glu Cys Pro Leu Leu Pro Pro Gln  
 45 35 40 45  
 ggc cgg ggt act gta ggg cag aca tta tgc cag tgg caa ctg gaa cgg  
 192  
 Gly Arg Gly Thr Val Gly Gln Thr Leu Cys Gln Trp Gln Leu Glu Arg  
 50 50 55 60  
 tac aca ggg atg tta ggg cga gtg gcc tgt ggt tcc ctg agc cag gta  
 240  
 Tyr Thr Gly Met Leu Gly Arg Val Ala Cys Gly Ser Leu Ser Gln Val  
 55 65 70 75 80

- 2 -

gat gag att tac cac gat gag tcc ctg ggg gtt cat ata aat att gcc  
 288  
 5 Asp Glu Ile Tyr His Asp Glu Ser Leu Gly Val His Ile Asn Ile Ala  
 85 90 95

ctc gtc cgc ttg atc atg gtt ggc tac cga cag tcc ctg agc ctg atc  
 336  
 10 Leu Val Arg Leu Ile Met Val Gly Tyr Arg Gln Ser Leu Ser Leu Ile  
 100 105 110

gag cgc ggg aac ccc tca cgc agc ctg gag cag gtg tgt cgc tgg gca  
 384  
 15 Glu Arg Gly Asn Pro Ser Arg Ser Leu Glu Gln Val Cys Arg Trp Ala  
 115 120 125

cac tcc cag cag cgc cag gac ccc agc cac gct gag cac cat gac cac  
 432  
 20 His Ser Gln Gln Arg Gln Asp Pro Ser His Ala Glu His His Asp His  
 130 135 140

gtt gtg ttc ctc acc cgg cag gac ttt ggg ccc tca ggg tat gca ccc  
 480  
 25 Val Val Phe Leu Thr Arg Gln Asp Phe Gly Pro Ser Gly Tyr Ala Pro  
 145 150 155 160

gtc act ggc atg tgt cac ccc ctg agg agc tgt gcc ctc aac cat gag  
 528  
 30 Val Thr Gly Met Cys His Pro Leu Arg Ser Cys Ala Leu Asn His Glu  
 165 170 175

gat ggc ttc tcc tca gcc ttc gtg ata gct cat gag acc ggc cac gtg  
 576  
 35 Asp Gly Phe Ser Ser Ala Phe Val Ile Ala His Glu Thr Gly His Val  
 180 185 190

ctc ggc atg gag cat gac ggt cag ggg aat ggc tgt gca gat gag acc  
 624  
 40 Leu Gly Met Glu His Asp Gly Gln Gly Asn Gly Cys Ala Asp Glu Thr  
 195 200 205

agc ctg ggc agc gtc atg gcg ccc ctg gtg cag gct gcc ttc cac cgc  
 672  
 45 Ser Leu Gly Ser Val Met Ala Pro Leu Val Gln Ala Ala Phe His Arg  
 210 215 220

ttc cat tgg tcc cgc tgc agc aag ctg gag ctc agc cgc tac ctc ccc  
 720  
 50 Phe His Trp Ser Arg Cys Ser Lys Leu Glu Leu Ser Arg Tyr Leu Pro  
 225 230 235 240

tcc tac gac tgc ctc ctc gat gac ccc ttt gat cct gcc tgg ccc cag  
 768  
 55 Ser Tyr Asp Cys Leu Leu Asp Asp Pro Phe Asp Pro Ala Trp Pro Gln  
 245 250 255

	ccc	cca	gag	ctg	cct	ggg	atc	aac	tac	tca	atg	gat	gag	cag	tgc	cgc	816
5	Pro	Pro	Glu	Leu	Pro	Gly	Ile	Asn	Tyr	Ser	Met	Asp	Glu	Gln	Cys	Arg	260
	ttt	gac	ttt	ggc	agt	ggc	tac	cag	acc	tgc	ttg	gca	ttc	agg	acc	ttt	864
10	Phe	Asp	Phe	Gly	Ser	Gly	Tyr	Gln	Thr	Cys	Leu	Ala	Phe	Arg	Thr	Phe	275
	gag	ccc	tgc	aag	cag	ctg	tgg	tgc	agc	cat	cct	gac	aac	ccg	tac	ttc	912
15	Glu	Pro	Cys	Lys	Gln	Leu	Trp	Cys	Ser	His	Pro	Asp	Asn	Pro	Tyr	Phe	290
	tgc	aag	acc	aag	aag	ggg	ccc	ccg	ctg	gat	ggg	act	gag	tgt	gca	ccc	960
20	Cys	Lys	Thr	Lys	Lys	Gly	Pro	Pro	Leu	Asp	Gly	Thr	Glu	Cys	Ala	Pro	305
	ggc	aag	tgg	tgc	ttc	aaa	ggt	cac	tgc	atc	tgg	aag	tcg	ccg	gag	cag	1008
25	Gly	Lys	Trp	Cys	Phe	Lys	Gly	His	Cys	Ile	Trp	Lys	Ser	Pro	Glu	Gln	325
	aca	tat	ggc	cag	gat	gga	ggc	tgg	agc	tcc	tgg	acc	aag	ttt	ggg	tca	1056
30	Thr	Tyr	Gly	Gln	Asp	Gly	Gly	Trp	Ser	Ser	Trp	Thr	Lys	Phe	Gly	Ser	340
	tgt	tcg	cgg	tca	tgt	ggg	ggc	ggg	gtg	cga	tcc	cgc	agc	cgg	agc	tgc	1104
35	Cys	Ser	Arg	Ser	Cys	Gly	Gly	Gly	Val	Arg	Ser	Arg	Ser	Arg	Ser	Cys	355
	aac	aac	ccc	tcc	cca	gcc	tat	gga	ggc	cgc	ctg	tgc	tta	ggg	ccc	atg	1152
40	Asn	Asn	Pro	Ser	Pro	Ala	Tyr	Gly	Gly	Arg	Leu	Cys	Leu	Gly	Pro	Met	370
	ttc	gag	tac	cag	gtc	tgc	aac	agc	gag	gag	tgc	cct	ggg	acc	tac	gag	1200
45	Phe	Glu	Tyr	Gln	Val	Cys	Asn	Ser	Glu	Glu	Cys	Pro	Gly	Thr	Tyr	Glu	385
	gac	ttc	cgg	gcc	cag	cag	tgt	gcc	aag	cgc	aac	tcc	tac	tat	gtg	cac	1248
50	Asp	Phe	Arg	Ala	Gln	Gln	Cys	Ala	Lys	Arg	Asn	Ser	Tyr	Tyr	Val	His	405
	cag	aat	gcc	aag	cac	agc	tgg	gtg	ccc	tac	gag	cct	gac	gat	gac	gcc	1296
55	Gln	Asn	Ala	Lys	His	Ser	Trp	Val	Pro	Tyr	Glu	Pro	Asp	Asp	Asp	Ala	420

- 4 -

cag aag tgt gag ctg atc tgc cag tcg gcg gac acg ggg gac gtg gtg  
 1344  
 5 Gln Lys Cys Glu Leu Ile Cys Gln Ser Ala Asp Thr Gly Asp Val Val  
 435 440 445

ttc atg aac cag gtg gtt cac gat ggg aca cgc tgc agc tac cgg gac  
 1392  
 10 Phe Met Asn Gln Val Val His Asp Gly Thr Arg Cys Ser Tyr Arg Asp  
 450 455 460

cca tac agc gtc tgt gcg cgt ggc gag tgt gtg cct gtc ggc tgt gac  
 1440  
 15 Pro Tyr Ser Val Cys Ala Arg Gly Glu Cys Val Pro Val Gly Cys Asp  
 465 470 475 480

aag gag gtg ggg tcc atg aag gcg gat gac aag tgt gga gtc tgc ggg  
 1488  
 20 Lys Glu Val Gly Ser Met Lys Ala Asp Asp Lys Cys Gly Val Cys Gly  
 485 490 495

ggt gac aac tcc cac tgc agg act gtg aag ggg acg ctg ggc aag gcc  
 1536  
 25 Gly Asp Asn Ser His Cys Arg Thr Val Lys Gly Thr Leu Gly Lys Ala  
 500 505 510

tcc aag cag gca gga gct ctc aag ctg gtg cag atc cca gca ggt gcc  
 1584  
 30 Ser Lys Gln Ala Gly Ala Leu Lys Leu Val Gln Ile Pro Ala Gly Ala  
 515 520 525

agg cac atc cag att gag gca ctg gag aag tcc ccc cac cgc att gtg  
 1632  
 35 Arg His Ile Gln Ile Glu Ala Leu Glu Lys Ser Pro His Arg Ile Val  
 530 535 540

gtg aag aac cag gtc acc ggc agc ttc atc ctc aac ccc aag ggc aag  
 1680  
 40 Val Lys Asn Gln Val Thr Gly Ser Phe Ile Leu Asn Pro Lys Gly Lys  
 545 550 555 560

gaa gcc aca agc cgg acc ttc acc gcc atg ggc ctg gag tgg gag gat  
 1728  
 45 Glu Ala Thr Ser Arg Thr Phe Thr Ala Met Gly Leu Glu Trp Glu Asp  
 565 570 575

gcg gtg gag gat gcc aag gaa agc ctc aag acc agc ggg ccc ctg cct  
 1776  
 50 Ala Val Glu Asp Ala Lys Glu Ser Leu Lys Thr Ser Gly Pro Leu Pro  
 580 585 590

gaa gcc att gcc atc ctg gct ctc ccc cca act gag ggt ggc ccc cgc  
 1824  
 55 Glu Ala Ile Ala Ile Leu Ala Leu Pro Pro Thr Glu Gly Gly Pro Arg  
 595 600 605

- 5 -

agc agc ctg gcc tac aag tac gtc atc cat gag gac ctg ctg ccc ctt  
 1872  
 5 Ser Ser Leu Ala Tyr Lys Tyr Val Ile His Glu Asp Leu Leu Pro Leu  
 610 615 620

atc ggg agc aac aat gtg ctc ctg gag gag atg gac acc tat gag tgg  
 1920  
 10 Ile Gly Ser Asn Asn Val Leu Leu Glu Glu Met Asp Thr Tyr Glu Trp  
 625 630 635 640

gcg ctc aag agc tgg gcc ccc tgc agc aag gcc tgt gga gga ggg atc  
 1968  
 15 Ala Leu Lys Ser Trp Ala Pro Cys Ser Lys Ala Cys Gly Gly Gly Ile  
 645 650 655

cag ttc acc aaa tac ggc tgc cgg cgc aga cga gac cac cac atg gtg  
 2016  
 20 Gln Phe Thr Lys Tyr Gly Cys Arg Arg Arg Arg Asp His His Met Val  
 660 665 670

cag cga cac ctg tgt gac cac aag aag agg ccc aag ccc atc cgc cgg  
 2064  
 25 Gln Arg His Leu Cys Asp His Lys Lys Arg Pro Lys Pro Ile Arg Arg  
 675 680 685

cgc tgc aac cag cac ccg tgc tct cag cct gtg tgg gtg acg gag gag  
 2112  
 30 Arg Cys Asn Gln His Pro Cys Ser Gln Pro Val Trp Val Thr Glu Glu  
 690 695 700

tgg ggt gcc tgc agc cgg agc tgt ggg aag ctg ggg gtg cag aca cgg  
 2160  
 35 Trp Gly Ala Cys Ser Arg Ser Cys Gly Lys Leu Gly Val Gln Thr Arg  
 705 710 715 720

ggg ata cag tgc ctg ctg ccc ctc tcc aat gga acc cac aag gtc atg  
 2208  
 40 Gly Ile Gln Cys Leu Leu Pro Leu Ser Asn Gly Thr His Lys Val Met  
 725 730 735

ccg gcc aaa gcc tgc gcc ggg gac cgg cct gag gcc cga cgg ccc tgt  
 2256  
 45 Pro Ala Lys Ala Cys Ala Gly Asp Arg Pro Glu Ala Arg Arg Pro Cys  
 740 745 750

ctc cga gtg ccc tgc cca gcc cag tgg agg ctg gga gcc tgg tcc cag  
 2304  
 50 Leu Arg Val Pro Cys Pro Ala Gln Trp Arg Leu Gly Ala Trp Ser Gln  
 755 760 765

aaa tac ttg ctg agc act agc tgt atg cca gac ctt gta cta aga atg  
 2352  
 55 Lys Tyr Leu Leu Ser Thr Ser Cys Met Pro Asp Leu Val Leu Arg Met  
 770 775 780

- 6 -

agg gaa cct agc atc aac cag aca gag ctg att ctt gcc ctc gtg cag  
2400  
Arg Glu Pro Ser Ile Asn Gln Thr Glu Leu Ile Leu Ala Leu Val Gln  
5 785 790 795 800

ccc aca gtc ttg tgc tct gcc acc tgt gga gag ggc atc cag cag cgg  
2448  
Pro Thr Val Leu Cys Ser Ala Thr Cys Gly Glu Gly Ile Gln Gln Arg  
10 805 810 815

cag gtg gtg tgc agg acc aac gcc aac agc ctc ggg cat tgc gag ggg  
2496  
Gln Val Val Cys Arg Thr Asn Ala Asn Ser Leu Gly His Cys Glu Gly  
15 820 825 830

gat agg cca gac act gtc cag gtc tgc agc ctg ccc gcc tgt gga gga  
2544  
Asp Arg Pro Asp Thr Val Gln Val Cys Ser Leu Pro Ala Cys Gly Gly  
20 835 840 845

aat cac cag aac tcc acg gtg agg gcc gat gtc tgg gaa ctt ggg acg  
2592  
Asn His Gln Asn Ser Thr Val Arg Ala Asp Val Trp Glu Leu Gly Thr  
25 850 855 860

cca gag ggg cag tgg gtg cca caa tct gaa ccc cta cat ccc att aac  
2640  
Pro Glu Gly Gln Trp Val Pro Gln Ser Glu Pro Leu His Pro Ile Asn  
30 865 870 875 880

aag ata tca tca acg gag ccc tgc acg gga gac agg tct gtc ttc tgc  
2688  
Lys Ile Ser Ser Thr Glu Pro Cys Thr Gly Asp Arg Ser Val Phe Cys  
35 885 890 895

cag atg gaa gtg ctc gat cgc tac tgc tcc att ccc ggc tac cac cgg  
2736  
Gln Met Glu Val Leu Asp Arg Tyr Cys Ser Ile Pro Gly Tyr His Arg  
40 900 905 910

ctc tgc tgt gtg tcc tgc atc aag aag gcc tcg ggc ccc aac cct ggc  
2784  
Leu Cys Cys Val Ser Cys Ile Lys Lys Ala Ser Gly Pro Asn Pro Gly  
45 915 920 925

cca gac cct ggc cca acc tca ctg ccc ccc ttc tcc act cct gga agc  
2832  
Pro Asp Pro Gly Pro Thr Ser Leu Pro Pro Phe Ser Thr Pro Gly Ser  
50 930 935 940

ccc tta cca gga ccc cag gac cct gca gat gct gca gag cct cct gga  
2880  
Pro Leu Pro Gly Pro Gln Asp Pro Ala Asp Ala Ala Glu Pro Pro Gly  
55 945 950 955 960

- 7 -

aag cca acg gga tca gag gac cat cag cat ggc cga gcc aca cag ctc  
 2928  
 5 Lys Pro Thr Gly Ser Glu Asp His Gln His Gly Arg Ala Thr Gln Leu  
 965 970 975

cca gga gct ctg gat aca agc tcc cca ggg acc cag cat ccc ttt gcc  
 2976  
 10 Pro Gly Ala Leu Asp Thr Ser Ser Pro Gly Thr Gln His Pro Phe Ala  
 980 985 990

cct gag aca cca atc cct gga gca tcc tgg agc atc tcc cct acc acc  
 3024  
 15 Pro Glu Thr Pro Ile Pro Gly Ala Ser Trp Ser Ile Ser Pro Thr Thr  
 995 1000 1005

ccc ggg ggg ctg cct tgg ggc tgg act cag aca cct acg cca gtc cct  
 3072  
 20 Pro Gly Gly Leu Pro Trp Gly Trp Thr Gln Thr Pro Thr Pro Val Pro  
 1010 1015 1020

gag gac aaa ggg caa cct gga gaa gac ctg aga cat ccc ggc acc agc  
 3120  
 25 Glu Asp Lys Gly Gln Pro Gly Glu Asp Leu Arg His Pro Gly Thr Ser  
 1025 1030 1035 1040

ctc cct gct gcc tcc ccg gtg aca  
 3144  
 30 Leu Pro Ala Ala Ser Pro Val Thr  
 1045

<210> 2  
 <211> 1048  
 35 <212> PRT  
 <213> Homo sapiens

<400> 2  
 Met His Ala Ala Ser Ala Ala His Ser Phe Val Leu Thr Ala Trp Leu  
 40 1 5 10 15  
 Pro Pro Trp Leu Ala Gln Lys Gly Ser Gln Val Leu Leu Leu Gln  
 20 25 30  
 45 Gly Pro Gly Val Cys Leu Leu Glu Glu Cys Pro Leu Leu Pro Pro Gln  
 35 40 45  
 Gly Arg Gly Thr Val Gly Gln Thr Leu Cys Gln Trp Gln Leu Glu Arg  
 50 55 60  
 50 Tyr Thr Gly Met Leu Gly Arg Val Ala Cys Gly Ser Leu Ser Gln Val  
 70 75 80  
 Asp Glu Ile Tyr His Asp Glu Ser Leu Gly Val His Ile Asn Ile Ala  
 55 85 90 95



- 8 -

Leu Val Arg Leu Ile Met Val Gly Tyr Arg Gln Ser Leu Ser Leu Ile  
 100 105 110  
 5 Glu Arg Gly Asn Pro Ser Arg Ser Leu Glu Gln Val Cys Arg Trp Ala  
 115 120 125  
 His Ser Gln Gln Arg Gln Asp Pro Ser His Ala Glu His His Asp His  
 130 135 140  
 10 Val Val Phe Leu Thr Arg Gln Asp Phe Gly Pro Ser Gly Tyr Ala Pro  
 145 150 155 160  
 Val Thr Gly Met Cys His Pro Leu Arg Ser Cys Ala Leu Asn His Glu  
 15 165 170 175  
 Asp Gly Phe Ser Ser Ala Phe Val Ile Ala His Glu Thr Gly His Val  
 180 185 190  
 20 Leu Gly Met Glu His Asp Gly Gln Gly Asn Gly Cys Ala Asp Glu Thr  
 195 200 205  
 Ser Leu Gly Ser Val Met Ala Pro Leu Val Gln Ala Ala Phe His Arg  
 210 215 220  
 25 Phe His Trp Ser Arg Cys Ser Lys Leu Glu Leu Ser Arg Tyr Leu Pro  
 225 230 235 240  
 Ser Tyr Asp Cys Leu Leu Asp Asp Pro Phe Asp Pro Ala Trp Pro Gln  
 30 245 250 255  
 Pro Pro Glu Leu Pro Gly Ile Asn Tyr Ser Met Asp Glu Gln Cys Arg  
 260 265 270  
 35 Phe Asp Phe Gly Ser Gly Tyr Gln Thr Cys Leu Ala Phe Arg Thr Phe  
 275 280 285  
 Glu Pro Cys Lys Gln Leu Trp Cys Ser His Pro Asp Asn Pro Tyr Phe  
 290 295 300  
 40 Cys Lys Thr Lys Lys Gly Pro Pro Leu Asp Gly Thr Glu Cys Ala Pro  
 305 310 315 320  
 Gly Lys Trp Cys Phe Lys Gly His Cys Ile Trp Lys Ser Pro Glu Gln  
 45 325 330 335  
 Thr Tyr Gly Gln Asp Gly Gly Trp Ser Ser Trp Thr Lys Phe Gly Ser  
 340 345 350  
 50 Cys Ser Arg Ser Cys Gly Gly Gly Val Arg Ser Arg Ser Arg Ser Cys  
 355 360 365  
 Asn Asn Pro Ser Pro Ala Tyr Gly Gly Arg Leu Cys Leu Gly Pro Met  
 370 375 380  
 55

- 9 -

	Phe	Glu	Tyr	Gln	Val	Cys	Asn	Ser	Glu	Glu	Cys	Pro	Gly	Thr	Tyr	Glu	
	385					390					395					400	
5	Asp	Phe	Arg	Ala	Gln	Gln	Cys	Ala	Lys	Arg	Asn	Ser	Tyr	Tyr	Val	His	
					405					410					415		
	Gln	Asn	Ala	Lys	His	Ser	Trp	Val	Pro	Tyr	Glu	Pro	Asp	Asp	Asp	Ala	
				420					425					430			
10	Gln	Lys	Cys	Glu	Leu	Ile	Cys	Gln	Ser	Ala	Asp	Thr	Gly	Asp	Val	Val	
			435					440					445				
	Phe	Met	Asn	Gln	Val	Val	His	Asp	Gly	Thr	Arg	Cys	Ser	Tyr	Arg	Asp	
15		450					455					460					
	Pro	Tyr	Ser	Val	Cys	Ala	Arg	Gly	Glu	Cys	Val	Pro	Val	Gly	Cys	Asp	
	465					470					475					480	
	Lys	Glu	Val	Gly	Ser	Met	Lys	Ala	Asp	Asp	Lys	Cys	Gly	Val	Cys	Gly	
20					485					490					495		
	Gly	Asp	Asn	Ser	His	Cys	Arg	Thr	Val	Lys	Gly	Thr	Leu	Gly	Lys	Ala	
				500					505					510			
25	Ser	Lys	Gln	Ala	Gly	Ala	Leu	Lys	Leu	Val	Gln	Ile	Pro	Ala	Gly	Ala	
			515					520					525				
	Arg	His	Ile	Gln	Ile	Glu	Ala	Leu	Glu	Lys	Ser	Pro	His	Arg	Ile	Val	
30		530					535					540					
	Val	Lys	Asn	Gln	Val	Thr	Gly	Ser	Phe	Ile	Leu	Asn	Pro	Lys	Gly	Lys	
	545					550					555					560	
	Glu	Ala	Thr	Ser	Arg	Thr	Phe	Thr	Ala	Met	Gly	Leu	Glu	Trp	Glu	Asp	
35					565					570					575		
	Ala	Val	Glu	Asp	Ala	Lys	Glu	Ser	Leu	Lys	Thr	Ser	Gly	Pro	Leu	Pro	
				580					585					590			
40	Glu	Ala	Ile	Ala	Ile	Leu	Ala	Leu	Pro	Pro	Thr	Glu	Gly	Gly	Pro	Arg	
			595					600					605				
	Ser	Ser	Leu	Ala	Tyr	Lys	Tyr	Val	Ile	His	Glu	Asp	Leu	Leu	Pro	Leu	
45		610					615					620					
	Ile	Gly	Ser	Asn	Asn	Val	Leu	Leu	Glu	Glu	Met	Asp	Thr	Tyr	Glu	Trp	
	625					630					635					640	
	Ala	Leu	Lys	Ser	Trp	Ala	Pro	Cys	Ser	Lys	Ala	Cys	Gly	Gly	Gly	Ile	
50					645					650					655		
	Gln	Phe	Thr	Lys	Tyr	Gly	Cys	Arg	Arg	Arg	Arg	Asp	His	His	Met	Val	
				660					665					670			
55	Gln	Arg	His	Leu	Cys	Asp	His	Lys	Lys	Arg	Pro	Lys	Pro	Ile	Arg	Arg	

- 10 -

	675					680					685					
	Arg	Cys	Asn	Gln	His	Pro	Cys	Ser	Gln	Pro	Val	Trp	Val	Thr	Glu	Glu
	690						695					700				
5	Trp	Gly	Ala	Cys	Ser	Arg	Ser	Cys	Gly	Lys	Leu	Gly	Val	Gln	Thr	Arg
	705					710					715					720
10	Gly	Ile	Gln	Cys	Leu	Leu	Pro	Leu	Ser	Asn	Gly	Thr	His	Lys	Val	Met
					725					730					735	
	Pro	Ala	Lys	Ala	Cys	Ala	Gly	Asp	Arg	Pro	Glu	Ala	Arg	Arg	Pro	Cys
				740					745					750		
15	Leu	Arg	Val	Pro	Cys	Pro	Ala	Gln	Trp	Arg	Leu	Gly	Ala	Trp	Ser	Gln
			755					760					765			
	Lys	Tyr	Leu	Leu	Ser	Thr	Ser	Cys	Met	Pro	Asp	Leu	Val	Leu	Arg	Met
	770						775					780				
20	Arg	Glu	Pro	Ser	Ile	Asn	Gln	Thr	Glu	Leu	Ile	Leu	Ala	Leu	Val	Gln
	785					790					795					800
	Pro	Thr	Val	Leu	Cys	Ser	Ala	Thr	Cys	Gly	Glu	Gly	Ile	Gln	Gln	Arg
25					805					810					815	
	Gln	Val	Val	Cys	Arg	Thr	Asn	Ala	Asn	Ser	Leu	Gly	His	Cys	Glu	Gly
				820					825					830		
30	Asp	Arg	Pro	Asp	Thr	Val	Gln	Val	Cys	Ser	Leu	Pro	Ala	Cys	Gly	Gly
			835					840					845			
	Asn	His	Gln	Asn	Ser	Thr	Val	Arg	Ala	Asp	Val	Trp	Glu	Leu	Gly	Thr
	850						855					860				
35	Pro	Glu	Gly	Gln	Trp	Val	Pro	Gln	Ser	Glu	Pro	Leu	His	Pro	Ile	Asn
	865					870					875					880
	Lys	Ile	Ser	Ser	Thr	Glu	Pro	Cys	Thr	Gly	Asp	Arg	Ser	Val	Phe	Cys
40					885					890					895	
	Gln	Met	Glu	Val	Leu	Asp	Arg	Tyr	Cys	Ser	Ile	Pro	Gly	Tyr	His	Arg
				900					905					910		
45	Leu	Cys	Cys	Val	Ser	Cys	Ile	Lys	Lys	Ala	Ser	Gly	Pro	Asn	Pro	Gly
			915					920					925			
	Pro	Asp	Pro	Gly	Pro	Thr	Ser	Leu	Pro	Pro	Phe	Ser	Thr	Pro	Gly	Ser
	930						935					940				
50	Pro	Leu	Pro	Gly	Pro	Gln	Asp	Pro	Ala	Asp	Ala	Ala	Glu	Pro	Pro	Gly
	945					950					955					960
	Lys	Pro	Thr	Gly	Ser	Glu	Asp	His	Gln	His	Gly	Arg	Ala	Thr	Gln	Leu
55					965					970					975	

- 11 -

Pro Gly Ala Leu Asp Thr Ser Ser Pro Gly Thr Gln His Pro Phe Ala  
                     980                                    985                                    990

5 Pro Glu Thr Pro Ile Pro Gly Ala Ser Trp Ser Ile Ser Pro Thr Thr  
                     995                                    1000                                    1005

Pro Gly Gly Leu Pro Trp Gly Trp Thr Gln Thr Pro Thr Pro Val Pro  
                     1010                                    1015                                    1020

10 Glu Asp Lys Gly Gln Pro Gly Glu Asp Leu Arg His Pro Gly Thr Ser  
                     025                                    1030                                    1035                                    1040

15 Leu Pro Ala Ala Ser Pro Val Thr  
                                     1045

## INTERNATIONAL SEARCH REPORT

In International Application No

PCT/EP 01/01525

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/57 C12N9/64 C12N15/62 A61K38/48 C12Q1/68  
C12N5/10 G01N33/50 C12Q1/37 C07K16/40 //A61P35/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N A61K G01N C12Q C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 00555 A (COLIGE ALAIN ;LAPIERE CHARLES (BE); PROCKOP DARWIN J (US)) 8 January 1998 (1998-01-08)	4
A	the whole document	6-11
X	--- DATABASE EM_HUM 'Online! EMBL; Accession number : AB002364, 13 February 1999 (1999-02-13) NAGASE, T. ET AL.: "Human mRNA for KIAA0366 gene" XP002169891 cited in the application abstract --- -/--	4

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*E\* earlier document but published on or after the international filing date

\*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*8\* document member of the same patent family

Date of the actual completion of the international search

18 June 2001

Date of mailing of the international search report

28/06/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Andres, S

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 01/01525

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>FROHME MARCUS ET AL: "Use of representational difference analysis and cDNA arrays for transcriptional profiling of tumor tissue." ANNALS OF THE NEW YORK ACADEMY OF SCIENCES, vol. 910, 2000, pages 85-105, XP001009700 ISBN: 1-57331-277-0 &amp; Conference on Colorectal Cancer New Aspects of Molecular Biology, Immunology and Clinical Applications; Berlin, Germany; October 08-09, 1999</p> <p>-----</p>	
A	<p>HURSKAINEN TIINA L ET AL: "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of zinc metalloproteases: General features and genomic distribution of the Adam-TS family" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 274, no. 36, 3 September 1999 (1999-09-03), pages 25555-25563, XP002158991 ISSN: 0021-9258 cited in the application</p> <p>-----</p>	

# INTERNATIONAL SEARCH REPORT

information on patent family members

In tional Application No

PCT/EP 01/01525

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9800555 A	08-01-1998	AU 3667397 A	21-01-1998
		CA 2260366 A	08-01-1998
		EP 0956354 A	17-11-1999
-----			

